



Master 1 DOP  
2021/2022

Analyse de Données  
Rapport de projet

## **« Analyse de la variation de la température de l'eau La Touques »**



**Encadré par :**  
Bruno DARDAILLON

**Réalisé par :**  
Thioune BASSIROU | 22008769  
Barry Mariama LAMARANA | 21910623  
Ndjim Mouhamadou MANSOUR | 21613317  
Yassine EL HARRAB | 21909693

22 Février 2022

# Plan

1. Introduction et contexte
2. Statistiques descriptives
3. Analyse en composantes indépendantes (ACI)
4. Analyse en composantes principales (ACP)
5. Conclusion

\* \* \*

## 1. Introduction et contexte

Dans le cadre d'un projet visant l'étude des cours d'eau et plus particulièrement la Touques (Fleuve de basse Normandie), 4 sondes placées en amont et aval le long du fleuve ont relevé les températures de l'eau le long du fleuve de la Touques. Ces sondes sont en général attachées aux racines des arbres de la sylvestre et immergées dans l'eau à une profondeur leur permettant d'enregistrer la température quelle que soit la saison.

La température de l'eau présente une variabilité saisonnière tandis que l'eau est généralement plus froide au fond (En raison de l'apport d'eau souterraine) et plus chaude à la surface (en raison de la température de l'air plus élevée).

Plusieurs activités ont une influence sur les évolutions de température d'un cours d'eau, en particulier on peut citer :

- Le rejet d'eau de refroidissement plus chaude des centrales thermiques,
- Le rejet d'eau des barrages,
- L'élimination des plantations riveraines qui ombragent et maintiennent les températures dans les cours d'eau,
- Une réduction ou une augmentation des niveaux d'eau en raison du captage ou du détournement de l'eau utilisée pour l'irrigation,
- L'ajout d'eau géothermique plus chaude etc.

Dans cette étude, nous nous intéressons à comprendre la variabilité de la température de l'eau (variable dépendante que nous cherchons à expliquer), en fonction de trois variables indépendantes ou explicatives présentes dans la base de données. Ces variables explicatives sont :

- La température de l'air
- Le volume de pluie
- Le type de sonde utilisé
- Le type de sonde

La base de données comprend 93 832 observations (sur une période allant de 2013 à 2018) et comprend 6 variables, présentées ci-après dans le tableau 1 :

Variable	Type de variable	Explication
<b>t</b>	Quantitative	temps, toutes les deux heures
<b>date</b>	Quantitative	date, en jour, mois et année
<b>Teau (<math>T_w</math>)</b>	Quantitative	température de l'eau
<b>id_sonde</b>	Qualitative (4 modalités)	sonde qui a relevé la température.
<b>Rainf.EOBS</b>	Quantitative	volume de pluie issu du site ECAD <sup>1</sup>
<b>Tair.EOBS (<math>T_a</math>)</b>	Quantitative	température de l'air issu du site ECAD

**Tableau 1** : Présentation des variables de la base de données par type de variable.

## 2. Statistiques descriptives

Les statistiques descriptives permettent d'avoir un aperçu général des données.

La température de l'eau varie selon le type de sonde qui prend la mesure. Le tableau 2 rend compte de la variabilité de la température (en particulier les valeurs extrêmes à savoir le minimum et le maximum). Le tableau présente également d'autres statistiques telles que la médiane, l'écart relatif, l'écart type, ainsi que le premier et troisième quartile).

Si l'on se concentre uniquement sur les valeurs extrêmes (min et max), nous pouvons constater que, toutes sondes confondues, la température minimum observée est de 1,897 et le maximum de température enregistré est de 21,135. Ces deux mesures (minimales et maximales) ont été enregistrées par la sonde 830. En revanche, la sonde 825 a enregistré une température minimale de 3,799 (soit la plus forte valeur de toutes les mesures prises par les sondes) et une température maximale de 17,809 (soit la plus faible valeur enregistrée pour le maximum de température).

Si l'on se réfère à une autre statistique importante, la moyenne, on peut noter que :

En moyenne, la température relevée par la sonde **825** est **2,14 %** moins élevée que celle prise par la sonde **827**. La température enregistrée, en moyenne par la sonde **827** est **1,61%** moins élevée que celle prise par la sonde **828**. Et enfin, la température moyenne relevée par cette dernière est **4,05%** moins élevée que celle prise par la sonde **830**. Ces résultats suggèrent donc que la température la plus basse est enregistrée par la sonde posée en amont.

	Type de sonde			
	Sonde 825	Sonde 827	Sonde 828	Sonde 830
Min	3.799	2.115	1.207	1,897
Max	17.809	18.545	20.208	21.135
Médiane	11.428	11.673	11.861	12.342
Moyenne	11.262	11.419	11.702	12.302
Ecart moyenne/médiane	1,45%	2,20%	1,30%	0,32%
Ecart-type	3.02244	3.294684	3.728789	3.695942
Q1	8,658	8.699	8.606	7,247
Q3	13,832	14.138	14.847	15.461

**Tableau 2** : Statistiques descriptives concernant la température enregistrée par type de sonde.

## 3. ACI : Analyse en Composantes Indépendantes

L'analyse en composantes indépendantes (ICA), quant à elle, est une méthode permettant d'identifier automatiquement les facteurs sous-jacents dans un ensemble de données donné (Stone, 2002). Il s'agit d'une technique statistique qui a pour objectif de décomposer un signal aléatoire multivariée en une combinaison linéaire de signaux indépendants.

En utilisant *fastICA*, nous avons pu extraire 3 composantes parmi les 4 qui soient les plus pertinentes.

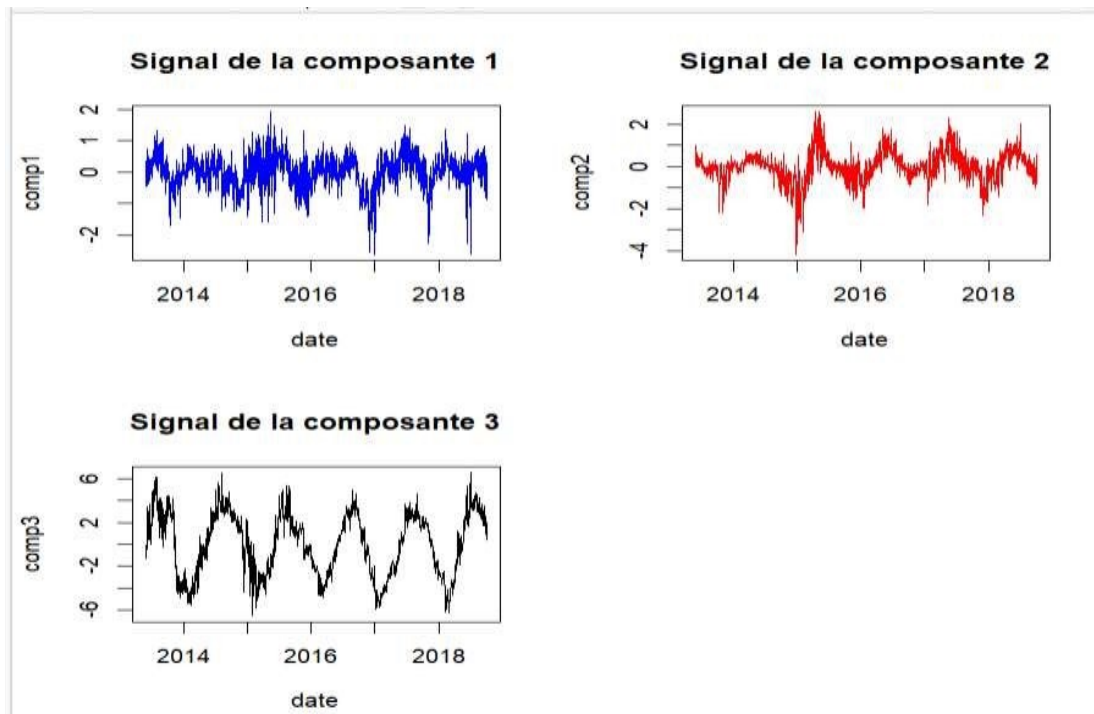


Figure 1 : Analyse ACI

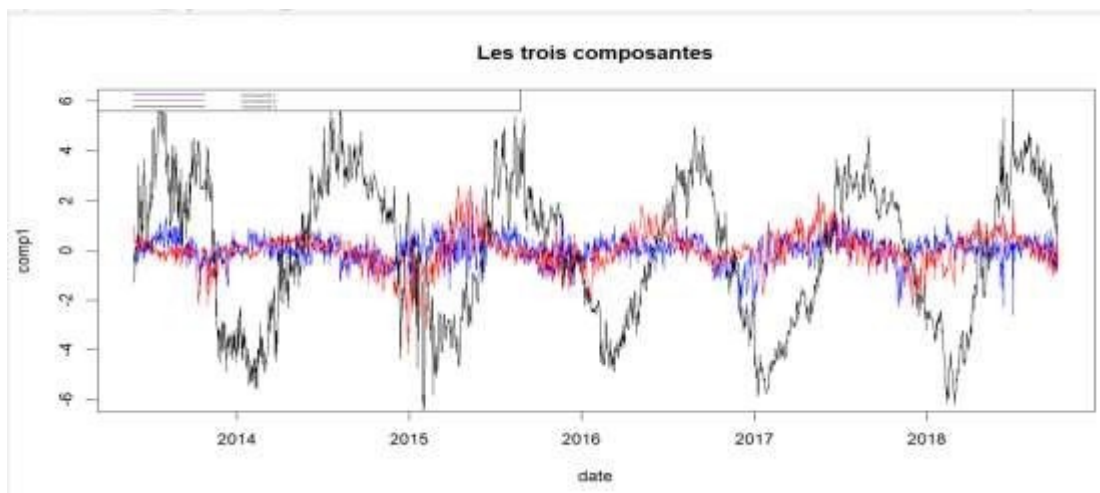


Figure 2 : Superposition des 3 analyses ACI

On observe qu'il existe une saisonnalité des composantes (figure 1). Nous pouvons constater que C2 semble décliner le plus par rapport à C1, prenant des valeurs extrêmes en 2015. Une des raisons pouvant expliquer cette observation est peut-être l'avènement de plusieurs variations météorologiques brusques.

Le déphasage des courbes (Figure 2) semble montrer que le signal d'une composante variant en premier pourrait être celui de la température de l'air (courbe bleu) et celui qui le suit (courbe rouge) serait celle de l'eau. Ceci semble cohérent étant donné que l'air se réchauffe relativement vite comparé à l'eau.

#### 4. ACP : Analyse en Composantes Principales

L'analyse en composantes principales (ACP) est une technique de la statistique multivariée. Elle consiste à transformer des variables liées entre elles ou corrélées en nouvelles variables décorréées les unes des autres. En effet, l'objectif de l'ACP est d'extraire les informations importantes de la base de données et de les représenter sous la forme d'un ensemble de nouvelles variables orthogonales. Ces dernières sont appelées composantes principales (Abdi and Williams, 2010). Grâce à la projection mathématique, l'ensemble de données

d'origine (comprenant de nombreuses variables) est réduit à un nombre plus faible de variables (Richardson, 2009). Les composantes principales concentrent l'essentiel de l'information contenue dans les données de départ. L'ACP est devenue l'un des outils les plus utiles pour la modélisation et de la visualisation des données (Vidal et al., 2016).

La mise en place de l'ACP nous permet de vérifier la corrélation existante entre les variables et de découvrir puis d'expliquer assez nettement nos composantes extraites de l'ACI.

Ainsi, une deuxième ACI avec *fastICA* nous a permis de décomposer en deux composantes le signal **D** (Tw-Ta) qui signifie la différence entre la température de l'eau et celle de l'air. Ces deux composantes que nous avons nommées **C1** et **C2** vont intervenir en tant que variables actives parmi d'autres variables, c'est-à-dire participeront à la fixation des axes de notre ACP.

Le diagramme des éboulis nous a permis de savoir qu'il serait beaucoup plus intéressant de travailler sur les deux premiers axes (Dim 1 et Dim 2) qui représentent à eux deux 92,6% de l'inertie totale. Ce qui est largement suffisant pour confirmer nos hypothèses précédemment établies.

L'ACP nous permet d'exploiter jusqu'à 5 axes d'interprétation. Mais nous allons nous concentrer que sur deux axes parmi les quatre dont les valeurs absolues sont supérieures à 1. On a donc appliqué le critère de **Kaiser**.

Ainsi, l'ACP proposée (**Figure 3**) résume **92 %** de l'information. Celle-ci est portée par l'axe 1 à **71 %** et par l'axe 2 à **21%**. Il y a 7 variables qui sont utilisées : **5 actives** (C1/C2/Tw/Ta/D) et **2 supplémentaires** (PE/P), Les variables actives participent à la fixation des axes et celles supplémentaires ne contribuent pas à la fixation des axes mais permettent de mieux interpréter notre graphique.

Deux tendances semblent se dégager : la dimension une (**Dim1**) est portée par **Tw, Ta et D**(Tw-Ta) à **74%** tandis que la dimension deux (**Dim2**) est fixée par **C1 et C2** à **95 %**. Ces valeurs nous proviennent du **tableau des contributions** issues de l'ACP.

Certaines variables apparaissent très liées comme Ta et Tw désignant respectivement la température de l'air et de l'eau, la **matrice des corrélations** vient confirmer ce rapprochement car les deux variables ont une corrélation de **0,91**.

Il apparaît bien une corrélation nulle entre les deux composantes C1 et C2 provenant de l'ACI donc indépendantes et dont les faisceaux sont parfaitement orthogonaux dès lors que les deux variables sont bien représentées dans le plan (1,2) à **0,90** et **0,99** respectivement pour C1 et C2 d'après le **tableau des cosinus carrés** désignant les qualités de représentation.

Les Ta et la composante C1 sont opposées. Ces deux variables sont fortement corrélées mais négativement (**-0,78** d'après la matrice des corrélations). Cette négativité traduit le fait que lorsque l'une augmente, l'autre diminue.

Les variables PE et P pourraient ne pas être assez bien représentées car très proches du centre du cercle.

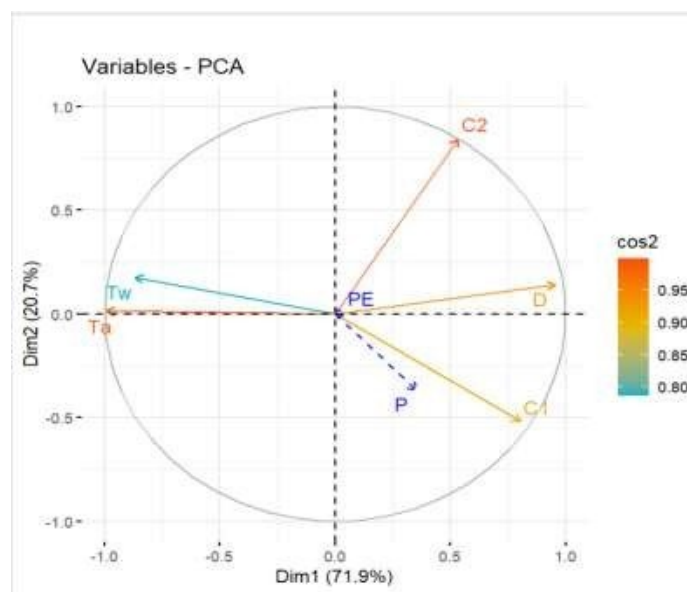


Figure 3 : Analyse ACP

## 5. Conclusion

La température de l'eau des principaux fleuves européens a augmenté de 1 à 3 °C au cours du siècle dernier. En particulier, plusieurs séries chronologiques montrent une augmentation des températures des lacs et des rivières dans toute l'Europe depuis le début des années 1900.

En particulier, une des variables qui a une influence sur l'évolution de la température de l'eau de surface des lacs et des rivières est la température de l'air. La température de l'eau devrait encore augmenter avec les augmentations prévues de la température de l'air. Les incidences de l'augmentation de la température de l'eau sont nombreuses. En particulier, l'augmentation de la température de l'eau peut entraîner des changements marqués dans la composition des espèces et le fonctionnement des écosystèmes aquatiques.

Dans cette étude, nous nous sommes basés sur un dataset comprenant plusieurs milliers d'observations sur la température de l'eau de la Touques entre 2013 et 2018. L'objectif de cette étude est de comprendre dans quelle mesure certains facteurs tel que celui que l'on vient de citer (température de l'air) et d'autres facteurs (type de sonde, pluviométrie) peuvent expliquer la variation de la température de l'eau. Afin d'y répondre nous avons utilisé deux techniques principales : l'ACI et l'ACP. Les principaux résultats de cette étude montrent que les phénomènes météorologiques influent fortement sur la température de l'eau.

## Références

- Abdi, H., Williams, L.J., 2010. Principal component analysis: Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–459. <https://doi.org/10.1002/wics.101>
- Richardson, M., 2009. Principal component analysis.
- Stone, J.V., 2002. Independent component analysis: an introduction. Trends Cogn. Sci. 6, 59–64. [https://doi.org/10.1016/S1364-6613\(00\)01813-1](https://doi.org/10.1016/S1364-6613(00)01813-1)
- Vidal, R., Ma, Y., Sastry, S.S., 2016. Principal Component Analysis, in: Generalized Principal Component Analysis, Interdisciplinary Applied Mathematics. Springer New York, New York, NY, pp. 25–62. [https://doi.org/10.1007/978-0-387-87811-9\\_2](https://doi.org/10.1007/978-0-387-87811-9_2)
- Van Vliet, M. T. H., Ludwig, F., Zwolsman, J. J. G., Weedon, G. P., & Kabat, P. (2011). Global river temperatures and sensitivity to atmospheric warming and changes in river flow. *Water Resources Research*, 47(2).

